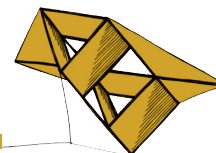


**STUDENTSKÁ  
VĚDECKÁ KONFERENCE 2019**

# Automatická syntéza pohybových dat pro vizualizaci znakového jazyka

Pavel Jedlička<sup>1</sup>

## 1 Úvod

Znakový jazyk (ZJ) je přirozený prostředek komunikace pro komunitu neslyšících. Ačkoliv je video považováno za přirozený způsob záznamu ZJ, pohyb se odehrává v 3D prostoru. Video je však pouze projekce prostorové scény do roviny. Technologie Motion Capture (česky lze přeložit jako Záznam pohybu) umožňuje zaznamenávat pohybová data v podobě 3D trajektorií a tím umožňuje zaznamenat pohyb vhodněji, viz Dilsizian et al. (2016). V současné době se k umělé syntéze ZJ používá nejčastěji pravidlově řízená syntéza. Ta však postrádá přirozenost lidského pohybu a proto je v praxi neslyšícími nepříliš přijímána, viz McDonald et al. (2016). Neuronové sítě využívající buňky LSTM (LSTM-NN) jsou navrženy pro práci s časovými posloupnostmi a jsou tedy vhodné pro práci s tímto typem dat, viz Martinez et al. (2017). Cílem této práce je využitím LSTM-NN generovat umělá pohybová data, která poslouží k animaci umělého znakového avatara. Tento avatar pak může posloužit například jako výstup automatického překladu.

## 2 Definice úlohy

Cílem této úlohy je prozkoumat možnost predikce pohybových dat pomocí LSTM-NN natrénované na relativně malém datasetu vysoce přesných pohybových dat. Vstupní dataset je složen z 334 promluv ve znakovém jazyce, kde každá promluva odpovídá jednomu zanku a je ukončena přechodem do tzv. klidové polohy (rest-pose), kterou řečník používá mezi promluvami. Data jsou definována jako trajektorie markerů umístěných na povrchu těla řečníka. Umístění markerů odpovídá topologii lidského těla a umožňuje reprezentaci polohy a orientaci jednotlivých kostí v prostoru. Celkový počet použitých markerů je 33 a jsou umístěny tak, že umožňují popsat pohyb trupu, ramen, paží a dlaní.

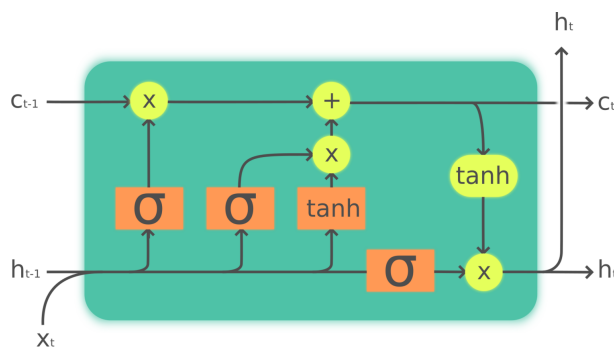
Data byla připravena tak, aby na vstupu byla sekvence o konstantní délce a délka výstupu je rovněž fixovaná. Toto omezení umožňuje přímé nasazení LSTM-NN pro řešení úlohy. LSTM síť byla implementována ve frameworku Keras. Tato implementace umožňuje vstup dimenze 2, jedná se o sekvenci vstupních vektorů  $v(t_0), v(t_1), \dots, v(t_N)$ . Výstupem je predikce vektoru (dimenze 1)  $v(t_{N+1})$ . Aby bylo možné predikovat výstup o délce větší než 1, výstup byl modifikován tak, že výstupní vektor je velikosti  $N \times M$ , kde  $M$  je délka výstupní sekvence.

### 2.1 LSTM-NN

LSTM buňka se skládá z buňky, vstupní, výstupní a paměťové brány (orig. input, output, forget gate). Buňka umožňuje pracovat s posloupnostmi dat tím, že zachovává minulé informace po určitou dobu. Tři brány pak regulují tok informace do a z buňky. Celá síť je pro danou úlohu

<sup>1</sup> student doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, e-mail: jedlicka@students.zcu.cz

složena z LSTM buňek v první vrstvě a fully-connected druhou vrstvou.



**Obrázek 1:** Buňka LSTM

### 3 Výsledky

Nejlepší dosažený výsledek dosahuje průměrné chyby trajektorie 6,96 mm pro všechny markery. Nejhorší výsledky predikce dosahují markery na zápěstí dominantní ruky znakovajícího s přesností průměrné chyby trajektorie 16,24 mm.

Testování optimálního nastavení parametrů i architektury sítě stále probíhá. Probíhají také experimenty s architekturami s hloubkou větší než 2 a také se způsobem generování výstupu s použitím LSTM-NN.

#### Poděkování

Příspěvek byl podpořen grantovým projektem číslo SGS-2019-027.

#### Literatura

- Naert, L., Larboulette, C., Gibet, S. *Coarticulation Analysis for Sign Language Synthesis*. Universal Access in Human–Computer Interaction. Designing Novel Interactions, Springer International Publishing, 2017.
- McDonald, J., Wolfe, R., Wilbur, R. B., Moncrief, R., Malaia, E., Fujimoto, S., Baowidan, S., Stec, J. *A new tool to facilitate prosodic analysis of motion capture data and a data-driven technique for the improvement of avatar motion*. 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).
- Dilsizian, M., Tang, Z., Metaxas, D., Huenerfauth, M., Neidle, C. *The Importance of 3D Motion Trajectories for Computer-based Sign Recognition* 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, The 10th International Conference on Language Resources and Evaluation (LREC 2016).
- Martinez, J., Black, M. J., Romero, J. *Human Motion Prediction Using Recurrent Neural Networks*, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)